DOCUMENT RESUME

ED 205 082 HE 014 108

AUTHOR Straumanis, Eric
TITLE Qualitative Evaluation and the Assessment of Program Quality. AIR Forum 1981 Paper.
PUB DATE May 81
NOTE 33p.: Paper presented at the Annual Forum of the Association for Institutional Research (21st, Minneapolis, MN, May 17-20, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment: *Educational Objectives: *Educational Quality: Higher Education: *Outcomes of Education: Prediction: Predictor Variables: *Program Evaluation: *Research Design
IDENTIFIERS *AIR Forum

ABSTRACT

The concept of quality is analyzed with special attention to its different senses in two contexts: qualitative evaluation and assessment of educational program quality. The view that evaluation of program quality cannot be validly expressed in quantitative terms and that qualitative evaluation methodologies have special relevance for the assessment of quality is criticized. An important step is to determine the central purpose or essential function of a program by expert judgment supported by good arguments. Statements of central purpose must be refined into statements of specific manifestations of quality. When the manifestations have been specified, it is possible to develop, through empirical research methods, indicators that predict the occurrence of the manifestations. In order to make a quality ranking, values of the indicated variables (which represent the manifestataions of quality) must be aggregated into a composite index of quality. If aggregation is not possible, profiling (i.e., talking of the various dimensions of value that fall within the scope of quality) is indicated. It is argued that valid quantitative and qualitative indicators can be developed if one has a model linking indicator with indicated variables. The distinctions that are made are designed to provide the conceptual framework that would underlie research designs for assessments of program quality. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED205082

09

QUALITATIVE EVALUATION

AND THE

ASSESSMENT OF PROGRAM QUALITY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Association for Institutional Research

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Eric Straumanis
Denison University
Granville, Ohio 43023
(614) 587-0810

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

NE 014 188

Presented at the Association for Institutional Research Annual Forum, Minneapolis/St. Paul, May 18, 1981.

Abstract

An analysis is made of the concept of quality, with special attention to its different senses in two contexts: qualitative evaluation and assessment of educational program quality. Arguments are presented against the views that evaluations of program quality cannot be validly expressed in quantitative terms and that qualitative evaluation methodologies have special relevance for the assessment of quality. It is argued that both valid quantitative and valid qualitative indicators of quality can be developed if one has a model linking indicator with indicated variables. The purpose of the paper is to establish distinctions which can serve as key components in a conceptual framework that would underlie research designs for assessments of program quality.

As retrenchment pressures in universities mount during this decade, assessments of the quality of educational programs will become increasingly important as a basis for deciding what is to be retained and what should be abandoned. But while the concept of quality is now often used by provosts and presidents in policy statements and directives, it is difficult to apply at the level of institutional research. The idea of quality seems to be bound up with the concept of value, but the latter, despite the great attention is has received in the last ten years, is still considered a problematic and elusive notion by most institutional researchers. To complicate matters, qualitative evaluation--a recently popular alternative to mainstream research methodology--is thought by some to have special relevance in the assessment of program quality. Others have maintained that "quality cannot be quantified". At least two symposia at the 1980 Association for Institutional Research (AIR) Forum dealt with the "qualitative dimensions of institutional research"; a conclusion of these sessions was that more clarity on this topic is needed. Hence, "Qualitative evaluation and the assessment of quality" marks out a methodologically important area in need of future work. This paper is intended as a contribution to that end.

The strategy of the paper is first to establish distinctions which can serve as key components in a conceptual framework that would underlie research designs for the assessment of educational program quality. These distinctions are then assembled to express a view of the assessment of quality which is

methodologically independent of qualitative evaluation. Finally, the merits of the qualitative methodologist's position are appraised, especially with respect to its challenge to the view developed in the earlier parts of the paper.

THE RELATIONSHIP BETWEEN QUALITY AND VALUE

Two Meanings of "Quality"

A convenient starting point for exploring the key distinctions relevant to the concept of quality is a consideration of some ordinary statements. Suppose that someone says, "Life in Plains, Georgia has a certain quality." The word "quality" in this case is being used non-committally or non-evaluatively. That is, the speaker is not recommending or appraising life in Plains. This has been called the descriptive meaning of "quality" (McCall, 1976). The point can perhaps be made clearer by paraphrasing the remark about Plains as follows: "Life in Plains, Georgia has certain qualities or qualitative features." Contrast this with the statement, "Atlanta has a quality cultural and intellectual climate." It is clear that the speaker is positively appraising or recommending life in Atlanta. It is this second, or evaluative meaning of "quality" which is of particular interest to program evaluators. Yet what has come to be known as qualitative evaluation (not to be confused with the evaluation of quality!) derives from the first, the descriptive meaning of "quality". This derivation will be explored in the second half of the paper. I shall now consider in some detail the evaluative use of "quality".

Quality as a Subcategory of Value

The value of something (let's follow tradition and call that something the value object) is its capacity to bring about or have certain desirable manifestations.[1] That capacity admits of determinations of degree. As a practical matter, it is very unlikely that anyone would care to assess the total or sum value of any, except perhaps the most simple, object. Even a relatively simple object such as a precious stone has multiple purposes or uses--aesthetic, industrial and investment, for example. These differing purposes or uses generate overlapping lists of desirable manifestations. Even in relatively simple cases the task of achieving compatible scaling is so formidable that trying to assess the object's total value is quite unfeasible. An educational program, considered as a value object, is much more complex than a precious stone, primarily because, unlike a stone, it is a set of processes.[2] It should be obvious that the assessment of the total value of as complex an object as an educational program is quite an unreachable goal.

Even if the concept of quality marks out a subcategory of the domain of value it might still not be worthwhile to attempt to determine the sum or total quality of an educational program. That is, does talking in terms of quality rather than value reduce the complexity enough to make assessment of quality a manageable task? Answering this question requires looking first at what we ordinarily mean by "quality".

Let's retreat to the case of precious stones because of its relative simplicity. What is presupposed when we judge that a stone possesses high quality? If we have its aesthetic purpose in mind, the judgment presupposes that the stone has the capacity to produce certain pleasing effects, or

that it has desirable manifestations. There is considerable consensus among lapidarists about the list of desirable aesthetic manifestations--color, clarity, brilliance, etc. Contrast this with the case where a stone is being evaluated for the purposes of investment. Although its quality as an aesthetic object would be an element in the investment decision, other considerations, possibly decisive ones, would enter into judging whether the stone has value for purposes of investment--such extrinsic factors as likely return on investment, which depends on world market trends, and so on. Notice that we don't speak naturally of quality in judgments of a stone's value for purposes of investment. (The sentence, "I made a quality investment today," seems strained, while "I invested in a quality stone" does not.) The reason we don't speak of quality is the presence of what we take to be extrinsic manifestations--factors that are not part of what is generally regarded to be the central function or essential purpose of precious stones--the aesthetic.[3]

If, however, a value object has desirable manifestations which are intrinsic to its central function or purpose, then we are justified in attributing quality to the object, i.e., a capacity to bring about or produce these manifestations. It is important to notice that quality, construed as a capacity (or dispositional property) is not directly inspectable or observable. To suggest that it is something like an intervening variable or a theoretical construct is not inaccurate; it is defined by reference to its effects or manifestations. I have used the word "manifestations" because I want to avoid the confusion that could ensue if I had chosen the more common term "outcomes", especially if one is tempted to take the apparently harmless short step from "outcomes" to "outputs". Though the

terminology of systems analysis (e.g., "input", "process", and "output" variables) can be very fruitful for certain research designs, considering outputs as the only type of manifestation of quality would be too restrictive. Although outputs would no doubt form part of the basis for nearly all attributions of quality, the central purposes of some value objects would entail manifestations of quality which on a systems view would be expressed as the values of process variables. For example, the central educational purpose of a department could be such that chairperson effectiveness, extent of faculty collegiality and degree of supportiveness in faculty-student relationships are manifestations (not indicators) of quality. But these were process variables in Bare (1980). Of course, since Bare was investigating department performance, not quality, it should not be surprising that performance was defined in terms of outcomes. But this difference should not affect the validity of my example here; "performance" and "manifestations of quality" no doubt have many items in common.

Because a precious stone is not a process, there is normally no need to try to predict its manifestations of quality. Nearly all of them can be simply "read off" by inspecting the object. (It would be a bit odd to perform a chemical analysis in order to predict that the stone will have a blue cast when viewed in daylight.) However, when the value object is an educational program which consists of many processes, then it is very useful to have indicators which make possible the prediction of desirable manifestations.

Manifestations, Indicators and Goals

At this point the reader might wonder whether the difference between manifestations and indicators is important. Don't both indicate the presence

or absence of quality? True, both do--in the sense that either can provide the warrant for a quality claim. But they differ in two important respects. First is a temporal difference, which has already been alluded to. That is, a good set of indicators makes possible an attribution of quality before the appearance of some of the manifestations of quality. If we have only the manifestations to go by, then the best we can do is to make after-the-fact attributions of quality. The second difference between manifestations and indicators is crucial to understanding the overall analysis set out in this paper. A manifestation of quality is conceptually tied to the agreed-upon central purpose of the type of value object under consideration. This relation is logical or necessary--it is not factual or empirical. For example, if part of the central purpose of the political science department is enabling students to classify certain events in specified ways and to frame researchable hypotheses, but it turns out that after completion of the department's program the students cannot do this, we would refuse to attribute high or even acceptable quality to that department. And the grounds for the refusal would be conceptual or logical--it would have been a misuse of the concept of quality to attribute it to the department, given the absence of relevant learning by students. The manifestations of quality are specifications of the central purpose of the value object--specifications conceptually or analytically derived from the statements of the central purpose. Or, to put it slightly differently, the manifestations of quality are intrinsic to the program's central function or essential purpose.

A good indicator predicts the occurrence of the manifestations of quality and hence bears a factual or empirical relation to the manifestations. Which indicators are reliable and valid is to be discovered by empirical

research. For example, whether student/faculty ratio or faculty research output or some combination of these or similar variables is a good indicator of some manifestation of quality, e.g. student achievement, is a factual question. Thus indicators indirectly indicate the presence of quality by predicting the appearance of the manifestations of quality.

The reader may have noticed that, rather than talking about goals of an educational program, I have chosen to talk in terms of central purposes or essential functions. My choice may appear incorrect in the light of some recommendations made by critics of American Council of Education (ACE) type reputational ratings of graduate department quality. These critics claim that "quality assessments must be referenced to departmental or institutional goals and objectives" (Lawrence and Green, 1980, p. 53). Although I agree that goal or objectives statements can provide important clues to the central purpose(s) of a program, they also tend to confuse manifestations with indicators of quality. This confusion is particularly likely in the present survey and research environment where disciplinary experts (or members of the "dominant coalition") are being asked not only to specify what I have called the manifestations of quality in their fields, but also to suggest indicators of quality. These two tasks are very different: the first calls for conceptual analysis or refinement of the agreed-upon central purposes for a particular field; the second is a request for speculation about empirical connections worth testing through research. But since these two tasks are seldom clearly distinguished, both candidate indicators and manifestations of quality are placed on departmental or program goal lists without any notice of their different methodological roles. That is, both manifestations and empirically unvalidated candidate indicators

of quality are referred to as "criteria of quality" and may be pursued as departmental objectives. This sort of mixup might not have adverse consequences for the conduct of the department or program, but it can be disastrous for the formulation of research designs for the assessment of quality. My use of the terms "function" and "purpose", instead of the more familiar terms "objective" and "goal", will not, of course, guarantee avoidance of this confusion. Mine is a terminological choice which tends to signal the need for researchers to extract the function or purpose of a program from goal statements and other kinds of relevant evidence. Finally, why the restriction to central purpose or essential function? Because, in my view, these terms identify that subset of the set of values exemplified by a program which we classify under the concept of quality. A comprehensive example, graphically presented, of the distinctions developed so far will be helpful at this point. (See Figure 1.)

EVALUATIVE DISAGREEMENTS AND THE PROBLEM OF SUBJECTIVITY

At this point the reader might be inclined to observe that, even if there is some benefit in having a conceptual scheme which distinguishes quality from its manifestations and indicators, and relates these to the central purposes of programs, little progress is possible unless there is a way of resolving value conflicts and answering charges of subjectivity.[4] The value conflicts are likely to arise when experts attempt to attain consensus on program central purposes and try to refine statements of purpose into statements of manifestations of quality. The charges of subjectivity are usually leveled at those whose "consensus position" has won out or has been voted in. My general reply--which I shall explain below--is that these are significant problems; the first can be largely resolved by

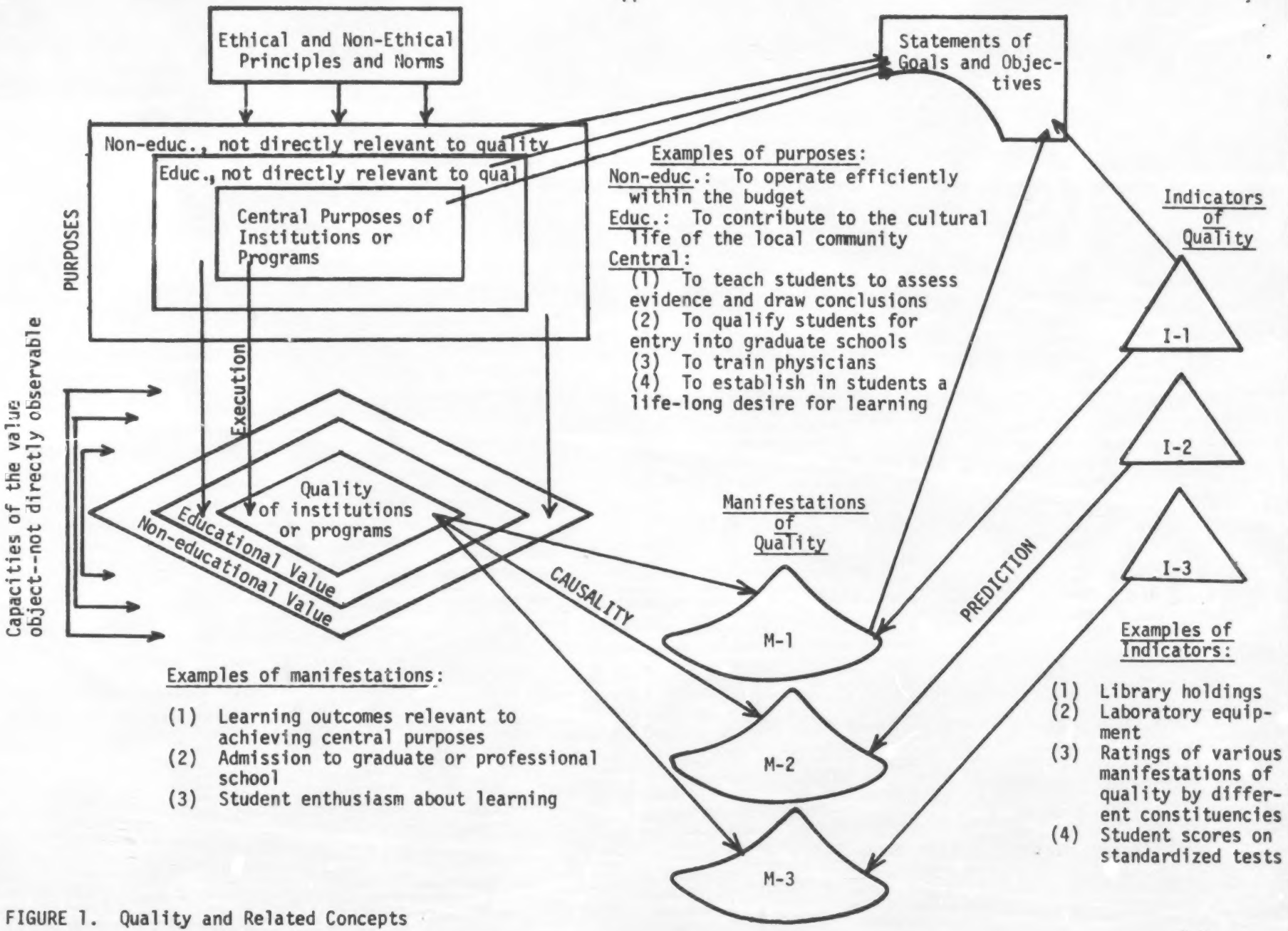
Ethical and Non-Ethical Principles and Norms
Statements of Goals and Objectives
Non-educ., not directly relevant to quality
Educ., not directly relevant to qual
Central Purposes of Institutions or Programs
PURPOSES
Examples of purposes:
Non-educ.: To operate efficiently within the budget
Educ.: To contribute to the cultural life of the local community
Central:
(1) To teach students to assess evidence and draw conclusions
(2) To qualify students for entry into graduate schools
(3) To train physicians
(4) To establish in students a life-long desire for learning
Indicators of Quality
I-1
I-2
I-3
Execution
Capacities of the value object--not directly observable
Quality of institutions or programs
Educational Value
Non-educational Value
Manifestations of Quality
CAUSALITY
PREDICTION
M-1
M-2
M-3
Examples of manifestations:
(1) Learning outcomes relevant to achieving central purposes
(2) Admission to graduate or professional school
(3) Student enthusiasm about learning
Examples of Indicators:
(1) Library holdings
(2) Laboratory equipment
(3) Ratings of various manifestations of quality by different constituencies
(4) Student scores on standardized tests

FIGURE 1. Quality and Related Concepts

using an approach different from the prevailing one; the second, though not fully resolvable, can be shown to have been much overrated as a problem.

The selection of central purposes for a program is not an empirical matter. It should be a matter of evaluative judgment backed by good reasons and arguments. But the prevailing approach by researchers seems to be to collect from experts or members of the "dominant coalition" intuitive or "unargued-for" judgments. Such judgments have the epistemological status of preferences rather than judgments with explicit warrant. It is not clear to me exactly why researchers deperd upon such survey data instead of collecting evaluative judgments which are supported by good reasons or arguments. Perhaps it is the desire to avoid what are thought to be "a priori evaluative standards" (Cameron, 1978, p. 606), or perhaps it is the sheer difficulty of collecting and handling data more complex than preference data. Whatever the reasons, one consequence of this approach is that the preference "votes" can easily outweigh opposing judgments supported by good reasons or sound arguments relating to a particular program's central purpose (and its refinements into particular manifestations of quality).

The enterprise of evaluative research still suffers from the widespread belief that value preferences, and even value judgments backed by good reasons, are inescapably subjective. As Scriven (1972) has shown, the word "subjective" is often misused. Just because a particular result or claim depends upon the judgments of persons (i.e. subjects) does not thereby make it subjective in the sense of <u>unreliable</u>. If certain precautions are observed, human beings make excellent evaluative instruments. But there <u>is</u> a degree of unreliability which attaches to the surveys described above--the averaging of the intuitive value preferences of experts in order to arrive at what

will function as justified central purposes of programs. As I have argued above, it is possible to do better than that by asking for good reasons and arguments as backing for the evaluative judgments. If we did the latter, then the problems of subjectivity (read "unreliability") would be largely eliminated.

Even if we improve our value data gathering approach as suggested above, evaluative disagreements will still occur. But nearly all of them could then be seen as subject to rational resolution. We would have avoided the error of immediately construing the disagreement as an "ultimate" clash of fundamental values, where "no position is better than any other". A conclusion that I draw in the next section is that the mistaken beliefs about subjectivity and its role in evaluative judgments might be contributing to the reluctance by researchers to aggregate the values of quality indicators.

AGGREGATION

It would be useful to develop measurement scales which enable scores on different indicators of quality to be aggregated to produce an overall index of quality for a particular type of educational program. If developed, the index would function much like a Quality of Life index which is a weighted average of individual scores on such social indicators as "security", "freedom from pollution", etc.[5] Although aggregation is not strictly required for the assessment of elements of value which are thought to constitute quality in a particular case, it is necessary if we wish our assessments to result in overall quality rankings of programs. Such rankings are needed for facilitating some kinds of policy decisions, for example, decisions about which of several programs should be allotted reduced resources.

It is interesting to notice that two researchers who have used very similar methodologies for arriving at quantitative indicators stop short of aggregation (forming a composite index) and rest with presenting profiles of indicator values. Mary Jo Clark says:

> The measures developed in this study provide reliable ways to obtain quantitative indicators of several previously identified quality-related program characteristics. But decisions about desirable levels of each characteristic . . . must still be made by experts . . . The crucial last step in the assessment of quality must and will continue to be judgments of value made by the men and women associated with the process (Clark, 1976, p. 102).

Similarly, Kim Cameron in his study of organizational effectiveness notes:

> It has been discovered that no institution operates effectively on all effectiveness dimensions, but that certain effectiveness profiles are developed in which particular dimensions are emphasized. No single profile is necessarily better than any other, since strategic constituencies, environmental domain, contextual factors, etc. help determine what combination is appropriate for the institution (Cameron, 1978, p. 625).

Some of the reasons for the reluctance to assign weights and to aggregate are clearly sound. It might be invalid to compare the aggregated scores of different types of institutions or programs. Also, a composite index of quality--even when controlled for program type and level--is bound to be politically unpopular because the absolute rankings that it makes

possible are often misused. The negative reactions to the results of the ACE reputational rankings of departments and the construction of invalid institutional rankings from the latter by the semi-professional educational press exemplify this problem. Moreover, if the sole purpose of the assessment is institutional or program self-improvement, then there is no compelling need for a composite index of quality.

Political unpopularity and the possibility of misuse should not inhibit efforts to develop composite indices of quality for various types of educational programs. Profiling the data and letting each institution or department decide "what combination is appropriate" might be taking the easy way out. To pick up a theme developed in the previous section, part of the reluctance to aggregate might stem from mistaken beliefs about the subjectivity of value judgments, beliefs which lead to the conclusion that the best we can do is to let each department, program or institution decide for itself what is relevant, and what weights are to be used, in an assessment of quality. My suggestion for making some headway against such subjectivity is to collect from experts judgmental data (and supporting arguments) regarding weights to be assigned to indicators and then have the data and arguments appraised by experts. This could remove one of the barriers which stands in the way to the development of valid composite indices of educational quality.

A BASIC CONCEPTUAL STRUCTURE FOR ASSESSING QUALITY

The following is a liberal summary (some new terms appear) of my position as it has been developed up to this point:

(1) The value of an object is its <u>capacity</u> to bring about certain

desirable effects. The concept of quality marks out a subcategory from the domain of the concept of value by picking out only those effects which are intrinsic to the object's central purpose or essential function.

(2) The quality of an educational program is its capacity to have certain desirable effects or manifestations. This capacity is a type-dependent complex dispositional property which is not directly inspectable, but whose presence must be inferred from either its manifestations or from indicators which predict the occurrence of the manifestations (sometimes called dimensions, features, characteristics, or criteria) of quality.

(3) What the central purpose or essential function of a program ought to be must be determined by expert evaluative judgment supported by good reasons or arguments, considerations which apply equally to the process of refinement of the statements of central purpose(s) into statements of specific manifestations of quality. The relation between manifestations and central purpose(s) is conceptual or non-empirical.

(4) When the manifestations of quality have been specified then it is possible to develop, through empirical research methods, indicators which predict the occurrence of the manifestations of quality. A somewhat technical way to say the same thing is that the values of the indicator variables make possible the prediction of the values of the indicated variables, providing one has a valid theoretical model which links the two types of variable.

(5) If we are to be able to attribute quality to the value object as a whole and to locate that object in an overall quality ranking, then we must aggregate the values of the indicated variables (which represent the manifestations of quality) into a composite index of quality. If we are

unable to aggregate, then we must be satisfied with "profiling", i.e., talking in terms of the various dimensions of value which fall within the scope of quality.

POSSIBLE CHALLENGES TO THE VIEW PRESENTED

The Holistic View

The recent literature on quality assessment contains several possible challenges to the position I have developed that I shall consider in the present section. These possible challenges are of particular interest because dealing with them requires a critical look at qualitative evaluation whose connection with the evaluation of quality I have promised to explore.

Stark and Lowther (1980), in reporting on the 1980 Association for Institutional Research (AIR) Forum panel discussion on the topic of measuring higher education quality, distinguish three views of the concept of quality. I shall not discuss the first two views which I consider to be not inconsistent with my own, but focus on the third, which is described as follows:[6]

> A third view, which carries the multidimensional view further, holds that the concept of quality is holistic. The ultimate measure is seen as the overall effect of the educational process on student growth and development. Quality is also seen as cumulative over time, implying that longitudinal rather than cross-sectional studies are the only valid technique to assess quality. Advocates of this position recognize that it is particularly hard to devise any uniform method for measuring the effect of education on individuals when (1) the population of students and the types of institutions are so

> varied; and (2) intellectual growth is not linear but varies in both rate and direction. The holistic approach calls for greater use of interview techniques, case studies, retrospective self-reports, existential analysis, and other qualitative methodologies focused primarily on assessing student growth but not totally ignoring the discernible effects of faculty and institutional settings on eventual outcomes (Stark and Lowther, 1980, p. 286).

Presumably what is meant by the first sentence in the above is not that the concept of quality is holistic, but that quality itself (i.e., the property of the value object) is holistic.[7] And the observation that the holistic view "carries the multidimensional view further" is, I believe, an understatement. I think "further" should be replaced with "to its logical extreme," though that way of putting it may not be fully satisfactory either. If I have correctly understood what is meant by "multidimensional view," then an example of that is the position developed by Bare (1980). Bare concludes from his research that departmental outcomes are predicted by relatively unique patterns of input and process variables; that department performance is unlikely to have any one set of predictors or indicators. To take this conclusion to its logical extreme would be to claim that no set of indicators can validly or meaningfully predict quality. At this extreme is located the holistic view which is a necessary part of qualitative evaluation methodology. The qualitative methodologist claims that essential meaning is lost when phenomena are "broken up" into the values of "predetermined" variables. I think it is this claim that is being referred to by the assertion, ". . . the ultimate measure is . . . the overall effect of the educational process on student growth . . ."

The other features of the holistic view, as represented by Stark and Lowther, are not incompatible with the standard, or reductionistic approaches to assessment. The need for longitudinal studies and for different criteria for value objects of different types and at different levels can be met without fundamental changes in reductionist methodology.

It is important to recognize that the claim that quality is holistic needs justification. One could not come to know by some sort of direct inspection that quality is holistic because, even on the qualitative methodologist's view, the nature and extent of quality must be inferred from its manifestations or effects. One candidate for a justification might seem to be the often cited principle that the whole is greater than the sum of its parts. But citing that does not get us very far because it is merely a definition of the concept of holism. An adequate justification would have to take the form of <u>evidence</u> that something essential is lost when we assess the quality of a program by considering no more than the quality-relevant values of its parts. It is beyond the scope of this paper to investigate whether there is such evidence and if so, to determine its adequacy. We can presume that the supporters of qualitative methodology are sensitive to the requirement to justify the principle of holism and believe that they can do so. For example, Patton (1980, p.75) holds that, "The state of the art of social science measurement is such that a number of desirable outcome measures still elude precise measurement. Where outcome measurements have not been carefully developed, it is more appropriate to gather descriptive information . . ."

The Connection Between Qualitative Evaluation and Quality

The strategy of qualitative research consists in a number of different methods which have most recently been discussed in detail by Patton (1980). As I explained at the outset of this paper, the term "qualitative evaluation" is not a synonym for "evaluation of quality", but stems from the descriptive meaning of "quality", not the evaluative meaning of that word. Indeed it is fine-grained description that is the hallmark of the qualitative approach for it expresses the methodologist's commitment to look very carefully at the phenomena--without any prior conception of what is to be collected and what can be allowed to "fall between the categories".

In describing procedures for the assessment of quality, Patton appears to have collapsed the distinction between the descriptive and the evaluative meanings of "quality". I shall quote him at some length in order to illustrate the subtle equivocation on the meaning of "quality".

> . . . to find out what it means to a student to have read a certain number of books is an issue of quality. How those books affected the student personally and intellectually is a question of quality. In contrast to counting the correct number of words spelled, the quality issue focuses on what spelling means to the student . . .
>
> . . . . . .
>
> Quality has to do with nuance, with detail, with the subtle and unique things that make a difference beyond the points on a standardized scale . . . .
>
> . . . . . .
>
> Failure to find statistically significant differences . . . on the outcome measure does not mean that there are no

> important differences . . . The differences may simply be qualitative rather than quantitative. A carpenter is reported to have explained this point to William James. The carpenter . . . observed: "There is very little difference between one man and another; but what little there is, is very important." Those differences are differences of quality (Patton, 1980, pp. 73-74).

Because he does not clearly distinguish (1) the qualitative (descriptive meaning of "quality") from (2) the valuable (evaluative meaning of "quality"), Patton is led to the strong, but unsupported, conclusion that educational quality cannot be meaningfully assessed unless one captures that which "falls between points on a standardized scale" in conventional, reductionistic research schemes. (I shall subsequently refer to this as the central claim of qualitative methodology.) In fact, his words seem to suggest an even stronger thesis which I hesitate to attribute to him: that only that which "falls between . . ." is relevant in the assessment of educational quality.

Theoretical Alternatives

I think that there is a way that Patton could provide theoretical support for the central claim, although doing so would not eliminate the need for justifying the qualitative approach but only move the justification back one level.[8] Nevertheless, looking at what the available theoretical moves are will prove fruitful. To explain this I will distinguish the following expressions:

(1) The whole is always greater than the sum of its parts.

(2) The whole is sometimes greater than the sum of its parts.

(3) The whole is never greater than the sum of its parts.

The qualitative methodologist can choose either (1) or (2) as the governing holistic principle. The reductionist has no choice but to accept (3) as the governing reductionist principle. Clearly, (1) is a stronger claim than (2).

If the qualitative methodologist can justify (1) for the domain of social or educational research, then it follows that quality can never be fully or validly assessed by reductionistic methods; qualitative methods must be used instead, though there is no guarantee that they will enable the researcher to attain the goal of fully describing quality. Because with (1) as the governing principle, the manifestations of quality cannot bear any specifiable, systematic relations to quality (there cannot be any dimensions of quality because quality is assumed to be holistic), neither the reductionist nor the qualitative methodologist can ever be sure of having attained a sufficiently fine-grained set of descriptions.

If the qualitative methodologist chooses (2) and can justify it, then it will be important to establish what types of value objects must be, and which need not be assessed by qualitative methods. On this alternative it would be difficult to give a general answer to the question of what kind of property quality is--complex or holistic? The answer depends upon the type of value object at issue in a particular case.

If (3) is justified then the reductionist wins and using qualitative methodologies provides no advantages, even if we grant the qualitative methodologist's claim to have an equally rigorous method. Since that method can be cumbersome and expensive, it would not be a good idea to expend scarce resources on it. On the reductionist view quality is a multidimensional property.

In this paper I have been developing and defending a version of the reductionist position. It should be clear that it is not inconsistent with this position to employ non-quantitative methods whenever quantification is inappropriate.

Quality and Quantification

Another challenge to the view I have advanced in this paper is represented by the arguments of Olscamp (1978). In a provocative article entitled "Can Program Quality Be Quantified?" Olscamp gives a negative answer to the question in the title. He supports that answer by employing the notion of "languages of quality." He says:

> What I shall try to do is to give a few samples from what I call "languages of quality" for properties of several things. If I cam correct, none of these languages of quality can be translated into quantitative symbolism; that is to say, these languages describe properties that make the thing or person which possesses them one of high quality, but these quality-producing properties cannot be described by a language of numbers or units that has the same meanings as does the language of quality. . . . Hence, if the thing being described were an academic program, for example, the program quality could not be quantified (emphases added) (Olscamp, 1978, p. 505).

According to Olscamp, it is faculty who have mastered the language of quality in their particular fields; outsiders have difficulty understanding descriptions of quality in terms of such "languages". Olscamp's position

appears extreme until one reads further where he allows that:

> To say that the quality of a thing cannot be described quantitatively does not mean that the thing cannot be scored, graded, or tested for the presence or absence of that quality. . . . What we mean when we say that Jones is tolerant is not that he scores well on some test, but that he treats people in certain ways. So the scoring is not a description of the quality, but rather an indicator of its presence or absence.

It is clear that Olscamp has set up translation which preserves meaning equivalence as a necessary condition for anyone's being justified in claiming that quality can be quantitatively described. And we know that this tough condition can never be met by any set of quantitative indicators if it is true that "the whole is always greater than the sum of its parts". As we saw in the preceding section, the qualitative methodologist can in principle always insist that any set of indicators, no matter how fine-grained, lets essential meaning slip between the categories. A fundamental normative question, which I have decided not to attempt to answer in this paper, is: What are the reasonable limits to such insistence?

It is significant that Olscamp (1977) recommends the development of "qualitative indices" which can be used to establish "quality ratings" of educational programs. But he does not make clear whether establishing quality ratings is just "scoring quality" or whether such ratings, when established by the use of qualitative indices, could also amount to a description of quality. What he has said at one point suggests the latter

alternative: "To describe quality we are required to describe types or classes of things with which we are concerned and then to explain what we mean when we say that . . . examples among the classes or types are good, better, or best among them (Olscamp, 1977, p. 197)." It seems that Olscamp recognizes the usefulness of indicators and even allows that there might be some point in ratings. But it is qualitative indicators which enjoy favored status. In addition to whatever practical advantages that Olscamp might claim for them, there is, I believe, a familiar theoretical basis for Olscamp's having given them favored status.

Instead of talking in the substantive mode--"the sum of the parts is always less than the whole"--Olscamp has advanced what one might call the semantic or linguistic version of this, now familiar thesis, what I earlier called the strong form of the qualitative methodologist's principle. That is, Olscamp maintains that "the quantitative indicators of X" can never mean the same as "the quality of X". Of course, the thesis in its linguistic form is in need of just as much justification as it is in its substantive form.

I shall conclude by arguing that the qualitative methodologist's principle has less practical import for the evaluation of quality than the supporters of the principle apparently think it has. It is time to pose a question to the qualitative methodologist: Who seriously claims that, in order to make defensible decisions involving educational quality, it is necessary to use a method which in principle can succeed in completely describing quality?[9] I think the answer is, nobody with some basic understanding of higher education. If this answer is correct, then the qualitative methodologist must be seen as having set up the proverbial "straw man" and then proceeded to knock it down with ease.

The bulk of educational decision-making which involves considerations of quality can get on quite well with reliable and valid indicators of the degree of its presence or absence. I am not suggesting that we have attained this happy state of affairs. But I <u>am</u> saying that we should not set up virtually unattainable methodological standards such as the requirement to be able to completely or exhaustively describe quality.

A final speculative note. When in a particular case, we feel that the resource allocation decisions of legislators, trustees, or senior university administrators are wrong-headed, we should hesitate to conclude that these decision-makers have misunderstood educational quality, that they have wrongly assumed it to be fully describable by some set of quantitative indicators. Instead we should, perhaps, be somewhat cynical and allow that decision-makers often quite deliberately give non-educational values and/or educational values not central to educational quality, precedence over educational quality. That this has happened in a particular case one usually hears only from critics, because, of course, it is politically difficult for the decision-maker to assert that quality has been outweighed by other values.

NOTES

1. There is a difference between the value of something (e.g. an educational program) and the values of a person. The latter are norms and principles in accordance with which the person acts. The values of decision-makers affect what value they attribute to an educational program, especially that part of the value of something they consider its quality.
2. For the sake of convenience I shall, in nearly all instances, use the term "educational program". However, I intend my analysis to be relevant at least at three levels: individual, program or departmental, and institutional.
3. Consensus about what is the central purpose of a value object may shift over time. For example, it is conceivable, though unlikely, that precious stones could cease to have an aesthetic function and that only the industrial function would remain. If this were to happen, we should expect a shift in the list of manifestations of quality in the evaluation of precious stones.
4. Value conflicts can be of different kinds and can therefore occur at different levels. Two decision-makers, A and B, can have an ultimate disagreement of principle, i.e., A and B disagree about the norms which justify particular program purposes. Or, A and B, while apparently in agreement at the level of principles or general norms, can disagree about how such norms should be "cashed out" in terms of program purposes, especially central purposes which determine the list of manifestations of quality. Finally, there can be conflict at the last stage--the derivation of statements of manifestations from statements of central purpose(s).

All except ultimate disagreements can be shown to be the result of factual and/or conceptual errors and hence are correctable by rational means.

5. In order for a composite index of educational quality to be fully like a Quality of Life index it will have to be a time-series. Presumably that could be accomplished through appropriate research design; Lawrence and Green (1980, p. 54) have suggested that this be done, but for reasons other than the one just mentioned.

6. Assuming that I clearly understand the first two views, I see no fundamental incompatibility between them, i.e., it is not clear how, on the second view, the matrix which expresses the first view has to be seen as "reductionistic". It seems to me that both views are "reductionistic" in that they assume that quality is a complex, multidimensional property of the value object, be that an individual, a program, or an institution. As I see it, the significant difference between the two views is that the first does not explicitly provide for the possibility of certain complex relationships between what I have called indicator and indicated variables, relationships demonstrated, though described somewhat differently, in Bare (1980).

7. What could it mean to say that a concept is holistic? Presumably all concepts are "holistic" in the sense of being univocal (having a determinate meaning), unless they are in need of explication. For example, the concept of indictrination might be considered to be equivocal or "multidimensional" unless it is carefully explicated. But I doubt that Stark and Lowther meant to refer to this distinction.

8. The qualitative methodologist's ability to provide theoretical support for the central claim does not eliminate the requirement to distinguish the descriptive and the evaluative meanings of "quality".
9. I have used the qualifying words "in principle" to signal the fact that the use of qualitative indicators does not guarantee success in fully describing quality. The most that the qualitative methodologist can claim is that if any method can succeed, it has to be a qualitative method.

REFERENCES

Bare, A. C. The Study of Academic Department Performance. Research in Higher Education, 1980, 12, 3-22.

Cameron, K. Measuring Organizational Effectiveness in Institutions of Higher Education. Administrative Science Quarterly, 1978, 23, 604-632.

Clark, M. J. The Meaning of Quality in Graduate and Professional Education. In J. Katz and R. J. Hartnett (Eds.), Scholars in the Making: The Development of Professional Students. Cambridge, Mass.: Ballinger, 1976, pp. 85-104.

House, E. R. Evaluating with Validity. Beverly Hills: Sage Publications, 1980.

Land, K. C. Social Indicator Models: An Overview. In K. C. Land and S. Spilerman (Eds.), Social Indicator Models. New York: Russell Sage Foundation, 1975, pp. 5-35.

Lawrence, J. K. and Green, K. J. A Question of Quality: The Higher Education Ratings Game. Washington, D.C.: American Association for Higher Education, 1980.

McCall, S. Human Needs and the Quality of Life. In J. King-Farlow and W. R. Shea (Eds.), Values and the Quality of Life. New York: Science History Publications, 1976, pp. 6-23.

Munitz, B. A. and Wright, D. J. Institutional Approaches to Academic Program Evaluation. In E. C. Craven (Ed.), Academic Program Evaluation. San Francisco: Jossey-Bass Inc., 1980, pp. 21-42.

Olscamp, P. J. Quality, Quantity, and Accountability. Educational Record, 1977, 57, 196-201.

Olscamp, P. J. Can Program Quality Be Quantified? Journal of Higher Education, 1978, 49, 504-511.

Patton, M. Q. Qualitative Evaluation Methods. Beverly Hills: Sage Publications, 1980.

Rossi, R. J. and Gilmartin, K. J. Handbook of Social Indicators. New York: Garland STPM Press, 1980.

Scriven, M. Objectivity and Subjectivity in Educational Research. In L. G. Thomas (Ed.), Philosophical Redirection of Educational Research. Chicago: National Society for the Study of Education, 1972, pp. 94-142.

Stark, J. S. and Lowther, M. A. Measuring Higher Education Quality. Research in Higher Education, 1980, 13, 283-287.

Van Alstyne, C. Postsecondary Education Indicators: Can We Measure Our Value Imperatives--Access, Diversity, and Excellence? In Planning, Managing and and Financing in the 1980's. Boulder: National Center for Higher Education Management Systems, 1978, pp. 77-88.